iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: VIII Month of publication: August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30835

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Analysis of Machine Learning Approach for Object Detection

Tanushree T N [1], Lakshman Naika R [2]
[1]P G Scholar Department of Electronics and communication, UNIVERSITY B.D.T. COLLEGE OF Engineering, Davangere-577004, Karnataka, India.
[2] Associate Professor Department of Electronics and communication, UNIVERSITY B.D.T. COLLEGE OF Engineering, Davangere-577004, Karnataka, India.

Abstract: Object detection is one of the main and most difficult computer vision branches widely used to track instances of semantically objects of a certain class in the lives of individuals, such as security monitoring, autonomous drive, etc. The efficiency of object detectors has been significantly enhanced with the rapid growth of deep learning networks for detection tasks. The architecture suggested uses pre-trained networks such as ALEXNET and VGG-16 to identify specific artifacts utilizing a PASCAL VOC 2007 dataset in today's language. The 25 layers of ALEXNET and VGG-16 are 41. Two principal directions are explored: supervised learning and semi-monitored learning. The disadvantages of supervised learning approaches drive unattended pre-training to be explored. By studying strong representations in early layers, layers can be educated quicker and more effectively.
Keywords: Convolutional neural network, Dropout, Batch normalization, Alexnet , VGG16.

I. INTRODUCTION

The development of hardware devices and algorithms plays an increasingly critical role in perception models. Researching object recognition is also considered as a general approach to artificial intelligence study since humans have a remarkable ability in both low- and high-resolution images to recognize large varieties of objects, although objects may vary in various perspectives, sizes and scales, and colours. As information processing elements in the visual system were discovered by Hubel and Wiesel in the earliest 1950s, many important methods of object recognition inspired by the building of artificial neural networks (ANNs) were used heavily during the 1990s. But soon, ANNs were decreased by the growth of support vector machines for object recognition. In recent competitions to recognize objects in the form of deep neural networks ANNs were not successful until 2012.

A. Deep feed forward

The function f ($\boldsymbol{x}$; $\boldsymbol{\theta}$), is the Feed-forward network and x is a network entry and $\boldsymbol{\theta}$ reflects network parameters. The aim of training feed-forward networks is to have a prediction about the corresponding target $\hat{\boldsymbol{y}}$ = f ($\boldsymbol{x}$; $\boldsymbol{\theta}$) approximate. It is known as multi-layer perceptrons were the deep feed-forward networks. Fig.1 shows a completely connected network (FC) with two layers concealed. This network can be represented by three single-layer feed-forward networks (excluding the input layer) connected in a chain where each is described by a functions f (i). So the network is formed by $f\ (\boldsymbol{x}) = f\ (1) \circ f\ (2) \circ f\ (3)\ (\boldsymbol{x})$. Let us simply consider the details of the first hidden layer, and then it can easily expand to the whole network. Its output is f (1) ($\boldsymbol{x}$) = ($\boldsymbol{W1x}$ + $\boldsymbol{b1}$), where $\boldsymbol{W1}$ be the weight matrix of the first hidden layer, b1 is its bias, and function g be its activation function.

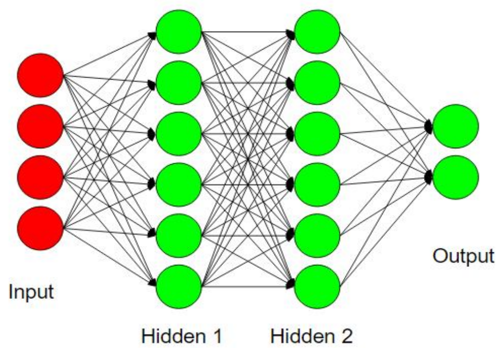

Fig. 1 Network with two hidden layers

II. LITERATURE SURVEY

In the BuyunLv et al., [1] a method integrated with vegetation index analysis was implemented with invariant transformation characteristics (SIFT), in 2014. The first is that the SIFT algorithm extracts feature points in two groups of points both in a vegetative area and in points in non-vegetation areas based on the vegetation index. The transition parameters are obtained with the lowest square rule after the removal of mixed points. Experimental results demonstrate to facilitate the proposed method achieve both high speed and good quality recording precision. Osman Orhan et al, [2] 2014 investigate the high temperature of the multi temporal land surface by using the vegetation temperature index, the vegetation index and vegetation status index. Pasher et al, [3] introduced linear woody characteristics tracking methods, 2016. For wood assessment, intersect line sampling is used. Curtis Chance et al, [4] 2016 the image correction method used to detect changes in the topographical surroundings. Rubina Parveen et al, [5] 2016, the photo was fitted with a partial differential equation (PDE). As metrics for vegetation identification, NDVI is used. For classification of vegetation contents, GLCM features are used.

Several studies have been conducted to automatically identify different targets. As building, aircraft, ships etc., to reduce human mistakes and save time and money. Due to their complex background, the differences in geometry, the need for topography and lighting for data acquisition and the range of objects for satellite images, it is nevertheless difficult. The detection task can be seen as the combination of two fundamental tasks to classify and determine the position of the objects in the images. To date, studies have focused on improving the two tasks individually or jointly. The first studies conducted most of the studies with uncontrolled methods using different attributes. For example, pan chromatic image detection was based on key transformation points for the scale invariant feature (SIFT) and the graph theorem. Instead a wavelet transform was used to classify the synthetic aperture radar (SAR) images. These unattended processes, however, produced typically effective results for specific structural types and were efficient for a limited number of artifacts. It is possible to establish high-performance different systems from more complicated scenes. The main reason for supervised learning's most successful results is that the process is performed in the training stage with manually marked samples. Structures before a convolutionary neural network (CNN) is used are normal.

III. PROPOSED METHOD

A. Convolutional neural network: CNN's are multiple layers. They are:

1) *Convolution layer*: this layer breaks the pictures into lesser parts to help detect features.
2) *Pooling Layers*: The pooling layers reduce the performance with non-linear sampling as well as lower the amount of parameter analysed through the network.
3) *Fully Connected Layers*: This layer multiples the entrance with a matrix of weight and introduces a function of bias.

B. Modern Activation Functions

ReLU is defined by the nonlinear function: g(u) =max(0, x). However, ReLU can “die” because the gradient is 0 when the unit is not active. Leaky ReLU attempts to alleviate this issue. Instead of being 0 when x < 0, leaky ReLU has a small negative slope (usually 0.01):

$$g(x) = \begin{cases} x, & x \geq 0 \\ 0.01x, & x < 0 \end{cases}$$

C. Dropout

Dropout inspired by the theory of sexual reproduction is designed to prevent the over-fitting problem in multiple nonlinear hidden layers. In the dropout layer, nodes are randomly removed with a keep-probability p in each training process. For example, if the keep-probability p is 0.8, only 20% randomly picked nodes can pass their values to next layer.

D. Batch Normalization

Whitened inputs have the benefits of speeding up training process, enabling higher learning rate and avoiding over-fitting problem. Ioffe S. et al. provides an approach to normalizing inputs to zero mean along with uniqueness variance. The approach is also known as batch normalization (BN). Batch normalization can also be considered as a network embedded before the activation functions.

$$\hat{\boldsymbol{x}} = BN(\boldsymbol{x}, X) = \boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]\sqrt{Cov(\boldsymbol{x})}$$

Where X is the set of the non-normalized inputs over the training data set $\{\boldsymbol{x}1, \boldsymbol{x}2, \ldots, \boldsymbol{x}N\}$, $\mathbb{E}[\boldsymbol{x}]$ is the expectation of the inputs, and $Cov(\boldsymbol{x})$ is the covariance matrix. The normalized $\hat{\boldsymbol{x}}$ has zero mean and unit variance.

E. Deep learning Pre-trained models

1) Alexnet

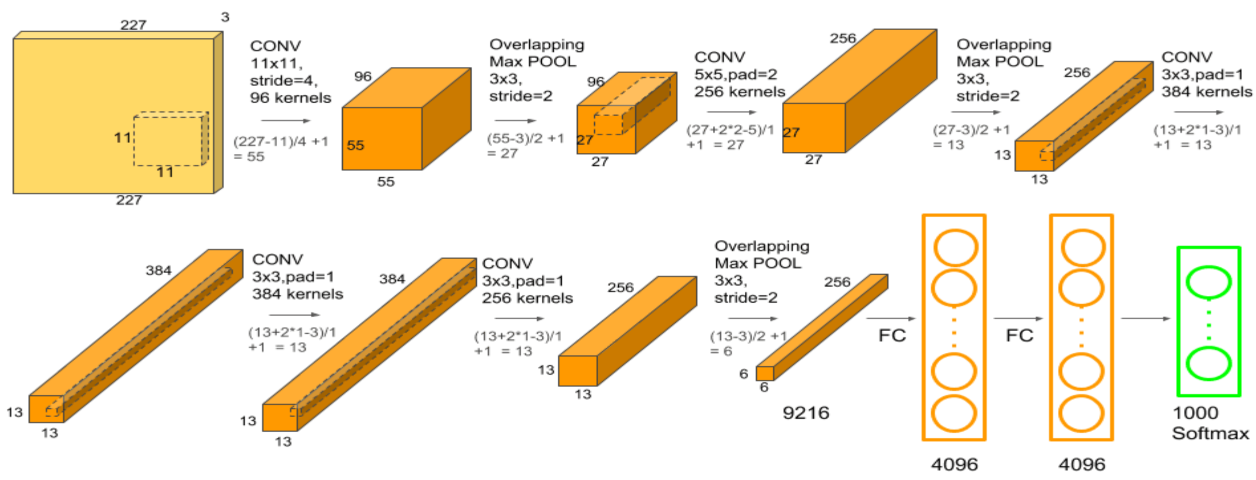

Figure 2: Architecture of the Alexnet.

a) 5 convolutional layer +2 FC layer+ softmax layer

b) Each convolutional layer- convolutional filters and a non linear activation function(RELU) ,3 layers have a max pooling layer

c) Input image size (227*227*3)

d) 60M parameters

e) 5 to 6 days to train on two GTX 580 3GB GPUs

f) Dropout feature

g) Data agumentation

2) VGG16

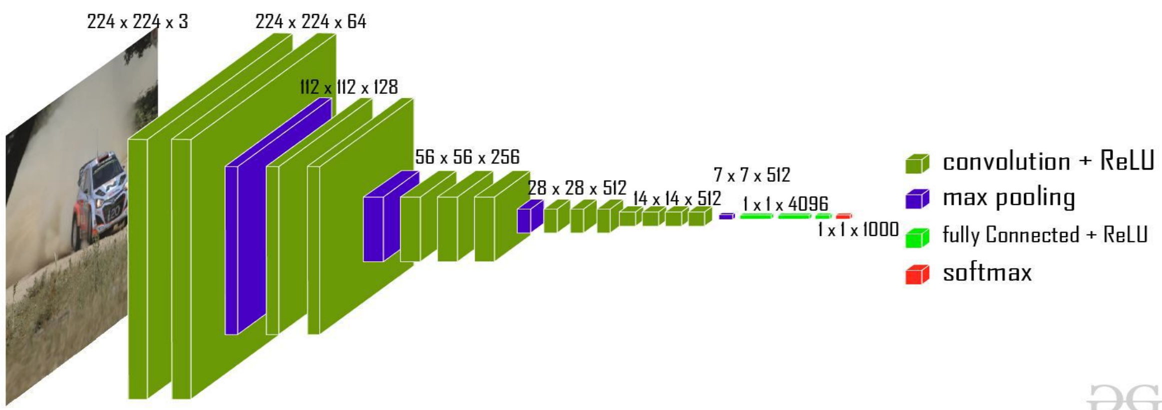

Figure 3: VGG-16 Architecture

a) VGG 16 was developed by Simonyan and zissernam by 2014

b) 16 convolutional layers

c) 138M parameters

d) Input image size(224*224*3)

e) Slow to train

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

IV. IMPLEMENTATION

The figure below shows the block diagram of this proposed framework that can detect objects through pre-trained, Google-released deep-learning networks.

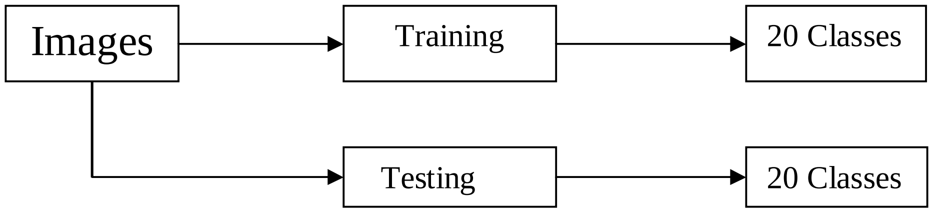

Figure 4: Block diagram.

The model is able to recognize the new data and predict the name of the probabilistic class the input image belongs.

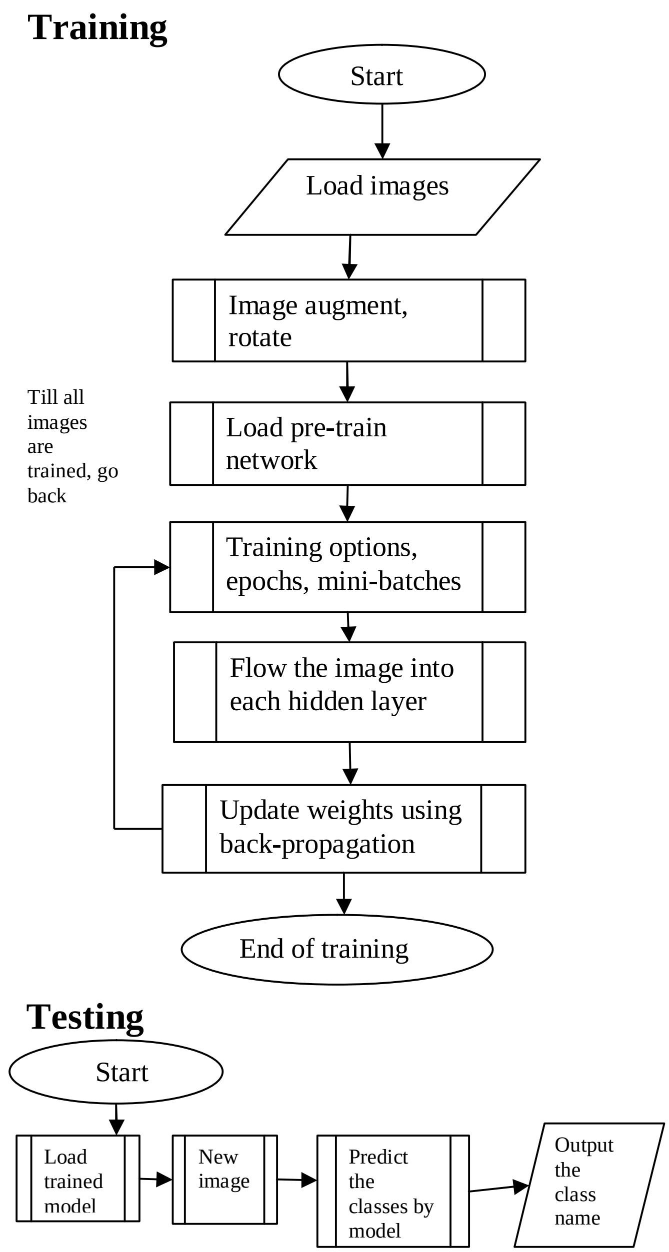

Figure 5: training and testing of the system

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

V. RESULT AND DISCUSSION

This chapter discuss about the snapshots obtained during the training and testing the model. The ALEXNET takes lesser training time compared to the VGG-16 network.

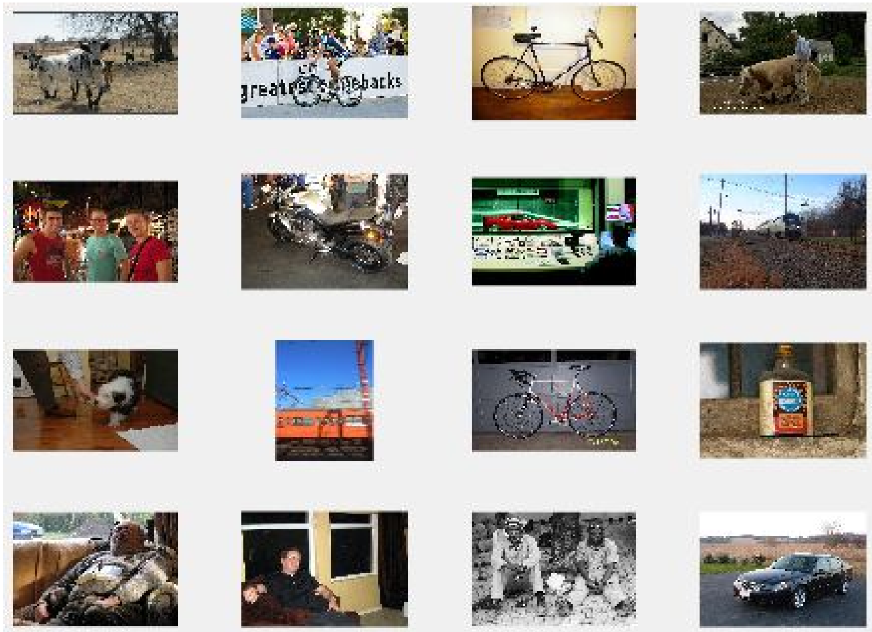

Figure 6: Snapshot of the Different classes.

A. Input Image

Figure 7: Final output of Alexnet

B. Input Image Rotated by 90 Degree

Figure 8: Final output of VGG16 net

VI. CONCLUSION AND FUTURE SCOPE

This thesis mainly focuses on the deep learning approaches to detect the objects in the image with the pre-trained networks such like ALEXNET and VGG-16. There accuracy parameters between these two networks are analyzed. The future can extend to the detection of video objects and the same work can be extended to detect stationary objects which may be used in military applications in order to monitor bombardment for long periods.

REFERENCES

[1] BuyunLv, Liaoying Zhao and Xiaorun Li, 2014, Remote sensing image registration using SIFT and vegetation index analysis, 7th International Congress on Image and Signal Processing, pp.575-579.
[2] Osman Orhan, SemihEkercin, and FilizDadaser-Celik, 2014, Use of Landsat land surface temperature and vegetation indices for monitoring drought in the salt lake basin area, Turkey, Hindawi Publishing Corporation, The Scientific World Journal, Vol.2014, pp.1-11.
[3] Pasher J., McGovern M., and Putinski V., 2016, Measuring and monitoring linear woody features in agricultural landscapes through earth observation data as an indicator of habitat availability, International Journal of Applied Earth Observation and Geoinformation, Vol.44, pp.113–123.
[4] Curtis M. Chance, TxominHermosilla, Nicholas C. Coops, Michael A. Wulder, and Joanne C. White, 2016, Effect of topographic correction on forest change detection using spectral trend analysis of Landsat pixel-based composites, International Journal of Applied Earth Observation and Geoinformation, Vol.44, pp.186–194.
[5] Rubina Parveen, Subhash Kulkarni and Mytri V.D., 2016, The Vegetation Extraction and Hierarchical Classification using an IRS-1C LISS III Image, International Journal of Computer Applications, Vol.155, No.9, pp.1-6.